S-157E
C.2.

FAMILY ORIENTED MEASURES OF EMPLOYMENT AND UNEMPLOYMENT

Prepared by

Statistics Canada
September 1983

I. INTRODUCTION

1. Traditionally, data on employment and unemployment have consisted of counts or estimates of the number of individuals in each of these states. In effect, each person has been viewed as an independent entity and little or no information has been published about the family context in which they live beyond, perhaps, their marital status. To some extent, this may have been due to the fact that in many countries data on unemployment were (and in some cases still are) obtained from unemployment insurance programs, and data on employment levels were (or are) derived from information supplied by employers. For a variety of reasons, both sources are ill-equipped to supply data on the family characteristics of the individuals concerned. However, the increasingly widespread use of household surveys (whether or not they are the source of the 'official' statistics), and the enrichment of labour force information in censuses of population, offer the potential of greatly enhancing the amount of data which relate the individual labour force participant to his or her family situation.

2. This note addresses selected aspects of the measurement of employment, unemployment, and labour market inactivity where the characteristics of the family are one of the dimensions of the measurement process. Most of the issues addressed concern the use of household sample surveys as the basis of estimation.

3. The document begins with a description of the two general types of family data which can be generated and suggests some of the types of applications to which each may be suited. It concludes with a review of some of the methodological problems associated with family oriented labour force statistics.

II. TYPES OF FAMILY DATA

4. There are two basic types of family oriented labour force statistics. The more familiar class of statistics consists of those estimates where the family is the unit of measurement. These data include estimates such as the number of two-parent families with both parents working, the number of families with one or more unemployed members, and the number of unattached individuals classified by their labour force status.

5. The second class of statistics is based on the individual as the unit of measurement. Here the family context in which the individual lives serves as a characteristic of the individual, analogous to the familiar demographic characteristics such as age and sex. Statistics in this group could include such measures as the labour force participation rate (activity rate) of women with preschool-age children, the number of unemployed single (lone) parents, the number of young labour force participants living with their parents, and so forth.

6. Obviously, statistics of both kinds are useful, with the choice between them being made according to the intended application.

III. CHOOSING THE APPROPRIATE TYPE OF FAMILY DATA

7. While one would not wish to denigrate the other benefits which accrue to the individual as a result of securing suitable employment, the income so derived is of paramount importance. Although the payment of wages or salary is made to the individual, the disposition of this income generally involves other members of the worker's family. (In Canada, for example, 87.2% of persons in the labour force live with at least one other person to whom they are related (1982 annual average)). Because of this sharing of earned income, the family is the most appropriate unit of analysis when measuring the economic welfare implications of labour force participation, employment, and unemployment.

8. For example, in the study of family income distributions, knowing how many family members are employed facilitates the determination of how many families enjoy 'acceptable' levels of income solely by virtue of having more than one earner, versus those who manage this by means of a single income. This capability, in turn, permits the assessment of the impact on family economic welfare of programs to create employment or, conversely, the consequences of restraint policies which may lead to reductions in employment. To take an obvious example, families which lose the income of a member whose contribution to the unit's economic welfare is minimal suffer less than families losing their sole source of work-related income.

9. Family statistics of this type also have more analytic applications such as the disaggregation of the growth in family income into the proportion due to increases in the incomes of individual family members, and the proportion associated with increases in the number of family members in employment.

10. When the unit of measurement is the individual, it is possible to relate these data to all of the standard measures such as the levels of employment and unemployment, as well as to the derived measures such as unemployment and participation rates. More specifically, the family characteristics of the individual can be used to disaggregate the conventional measures of levels and rates and in so doing make these conventional measures more useful and increase their 'explanatory power'.

11. As an example of the enhancement of 'explanatory power' (or in this instance 'predictive power'), one could cite the labour force participation rates (activity rates) of females, disaggregated by the presence or absence of preschool-age children, and the participation rates of older males living in families versus those living alone. Canadian data have shown that the rise in female participation rate can be largely attributed to the progressively increasing rates of one sub-set, i.e., women with preschool-age children, and that the decline in the overall participation rate of older males is substantially confined to declines in the rates for men who are not living with persons to whom they are related.

12. In terms of measuring economic welfare, it is perhaps best to use the family as the unit of measurement and to explain changes in economic welfare, or the determinants of family economic welfare, according to the number of persons who are employed or unemployed. However, the family is not the usual unit of measurement when reporting statistics on the level of employment or unemployment. Therefore, when there are advantages to reporting data with the individual as the unit of measurement (and such advantages may amount to nothing more than adherence to customary practice), it is useful to have the data disaggregated by family situation when attempting to convey the economic welfare implications of the statistics. To take an obvious example, a rise in the overall level of unemployment which is composed of an increase in the number of persons living in families with two or more members employed full-time has much different welfare implications than a rise of the same magnitude which can be largely attributed to persons living in large families with no member employed.

13. Family data with the individual as the unit of measurement can also be of considerable value for what might be described as analytic applications. To cite but one example, in times of increasing unemployment, a 'counter-cyclical' increase in the participation rate of certain demographic groups can be more accurately attributed to the so-called 'additional worker effect' if it can be shown that this increase is found among individuals who are members of families with no employed members. (Clearly, longitudinal family estimates are best suited to this type of application but a discussion of this type of data is beyond the scope of this paper.) Other applications of the same 'genre' are also possible such as explaining declines in post-secondary school enrolment in terms of whether or not the parents of the persons in question are employed or unemployed.

14. In addition to contributing to an understanding of the decision to participate or not in the labour force, the family dimension may contribute to an understanding of the type of participation. For example, it may well be found that persons electing to work part-time seem to do so because of the presence of other family members such as the very young and the very old who make demands on the non-working time of these persons.

IV. DATA GENERATION TECHNIQUES

15. Labour market statistics come primarily from two sources, that is, administrative sources and households surveys (including population censuses). No blanket statement can be make about administrative statistics which would universally reflect national practices. However, it would not be a serious distortion to say that most programs from which labour market oriented administrative statistics are drawn generate records which represent individual program participants, be they employed, unemployed, or outside of the labour force.

16. Under these circumstances, it is very difficult to derive family statistics, whether the unit of measurement is the individual or the family. The extent to which family statistics can be produced at all depends on the amount of family-related information contained on the records. Even where such information constitutes a portion of the information concerning individual program participants, it is not likely to be very extensive since it will, in all probability, only concern the aspects of the individual's family situation needed to administer the program from which the records are derived. Any expansion in the range of family data would require that the program administrators expand the amount of information collected from, or about, the program participants. In recognizing the logical priorities of program administrators, it can be appreciated why they would be less than enthusiastic about taking on such burdens merely to serve statistical ends.

17. In the case of household surveys and population censuses, the methodologies employed are generally well suited to the generation of both types of family statistics, even when the production of statistics on individuals has been the primary consideration in the development of the methodology. The necessary conditions include:

(i) All persons living in a dwelling be included in the survey.

(ii) A distinct and recognizable record is generated for each family member.

(iii) The survey records carry 'linkage keys' so that all members of a given household or family can be identified in the survey files.

(iv) The question set addressed to all respondents includes a question on the relationship (spouse, son, daughter, etc.) among family or household members.

18. Given that these conditions have been satisfied, the generation of the two types of family estimates described earlier can be achieved by following what might be described as a four-stage procedure:

(i) The first stage involves assembling the survey records into groups, each of which constitutes a family. (In the case of unattached individuals, such a 'group' consists of only one record.)

(ii) The second stage involves deriving the characteristics of the family from the group of individuals records. A simple count of the number of records establishes family size, and this in turn, can be readily disaggregated into the number of family members belonging to each age group. For family members of working age, reference to individual labour force statuses establishes the number who are employed, unemployed, or not in the labour force. At this stage, specific combinations can also be determined. For example, in husband/wife families, it is possible to identify cases where both the husband and wife are employed, where one is employed and the other unemployed, where neither is employed, and so on. Particular family formations can also be made a matter of record, such as the

family being 'headed' by one person (and whether that person is male or female), whether the family is composed of one, two, or three generations, etc. In short, a rather lengthy vector of family characteristics can be established with the length of the vector being only limited by the number of characteristics of the individual family members recorded by the survey.

(iii) This vector of family characteristics is then treated in two ways. Firstly, it is used as a new record in its own right. The assemblage of such records then constitutes the file from which estimates where the family is the unit of measurement are derived. Secondly, this vector is added to the original records representing each of the family members concerned.

(iv) The final stage consists of adding weights to both files so that the results of the sample can be converted into estimates of the characteristics of the entire population. Such weights are obviously required in the case of sample surveys but they may be necessary in the case of censuses as well, at least when certain census questions are addressed to only a portion of the population. (For example, in the 1981 Canadian Census of Population, questions on labour force status were asked in only one household in five.) Issues relating to the weighting of both types of files are addressed in the next section.

V. ESTIMATION PROBLEMS

19. Over the years, innumerable papers have been written on the topic of the 'harmonization' of data from different sources. These documents frequently deal with cases where inconsistencies have their origin in the use of different methodologies (e.g., establishment versus household surveys) and the associated differences in measurement errors.

20. In the case of family statistics, similar problems of consistency can arise even though the two types of family data come from the same source, i.e., a single household survey. Since both types of estimates are derived from the same basic set of records, these disparities arise from differences in the weighting procedures.

21. Simplifying somewhat, weighting schemes generally involve three stages. In the first stage, each record is assigned an initial weight which is the inverse of the sampling ratio. In the case of records derived from surveys based on samples of dwellings or households, the inverse of the sampling ratio will be the same whether the record refers to an individual, a family or a household. For example, in a survey covering one in every 200 households, the initial weight will be 200. The second stage involves a variety of adjustments, the most significant of which is generally an adjustment for non-response. Since, as a rule, entire households rather than individual household members are non-respondents, the non-response adjustment is applied equally to records for families and individuals. When weighting schemes are of the type just described, the weights are the same for all units of measurement (i.e., households, families and individuals) after the first two stages.

22. In some surveys, the estimation methodology consists solely of these two stages. In the present context, this has the very real advantage of preserving consistency among the estimates of the two types of family data. In other words, the estimated number of heads of families of size 'n' will be exactly equal to the number of families of size 'n'. Similarly, the number of employed married women with an employed spouse present will be precisely the same as the number of employed married men with an employed spouse present. This property derives from the fact that the family weight equals the weights associated with the records representing each and every member of that family. However, there are serious disadvantages to a weighting scheme of this nature. In particular, for a given sample, the sampling variance will be higher than that under weighting algorithms with more stages, and the estimates will be characterized by a downward bias due to undercoverage. (Undercoverage occurs when dwellings or persons in dwellings, who should be enumerated according to the survey's design, are not in fact covered. The result is that the de facto sampling ratio is lower than the design ratio and this is not compensated for in the weight.) This latter deficiency is particularly serious in the case of family related statistics because it has been found that specific age-sex groups are more prone than others to undercoverage, and small families and unattached individuals are more likely to go unenumerated than large families.

23. With a view to minimizing the sampling variance and adjusting for undercoverage, sample surveys frequently make use of a third stage of weighting which involves adjusting the results to 'control totals' derived from sources which are statistically independent of the survey.[1,2] In other words, the weights on the records are modified so that after modification, the sum of the weights is equal to the control totals. For the file of individuals, a number of age-sex specific control totals can be used, and for the file representing family units, the control totals can be disaggregated by family size.[3] The advantages gained by the use of this third weighting stage increase with the increase in the degree of disaggregation in the control totals up to a certain limit. (The nature of this limitation will not be detailed here.)

---

1 For an exposition of the effect of the third stage of weighting on sampling variance, see M. Lawes and M.P. Singh, Comparison of Some Ratio Estimators For Large Scale Household Surveys, presented at the 1980 Meeting of the Statistical Society of Canada.

2 Another benefit of this adjustment is that it tends to reduce the variation in the estimates from one observation to the next. This is a reflection of the absence of spurious variation in the control totals which is generally a deliberate property of such estimates. While this reduced variation in the survey estimates has no statistical advantages from a purely scientific point of view, estimates which exhibit 'reasonable' changes from one observation to the next (in terms of both magnitude and sign) are more likely to enjoy enhanced credibility in the eyes of the general data user community. The more frequent the observations, the more this property will be appreciated.

3 In some cases, rather than using family control totals to perform a third stage estimation, one simply takes the weight associated with a family member, after the third stage adjustment and uses that as a family weight. This can be seen as a partial solution to the undercoverage problem in the absence of a set of suitable family control totals. It does not, however, solve the inconsistency problem described beginning in paragraph 24.

24. Unfortunately, while this stage has a number of advantages for both estimates of the family characteristics of individuals and estimates relating to families, it is this stage which introduces the inconsistencies within selected elements of the set of family related estimates where the individual is the unit of measurement.

25. All of these inconsistencies have their origin in the fact that undercoverage varies by age and sex, and by family size.[4] The third stage in the weighting process therefore introduces differences in the weights of various members of a given family and results in a family weight which may differ from the weights associated with the family's members. This will result in such logical inconsistencies as the number of husbands and wives differing, the number of family 'heads' differing from the number of families, and total family income differing from total individual income.

26. These disparities will, as a rule, not be very large in a well designed and managed survey. When they are large, the implication is that the survey suffers from significant undercoverage. For sophisticated users of the data who understand the estimation methodology and the limitations of the estimates, the lack of consistency will not be disquieting. However, for many users, or possibly most, such discrepancies may be taken as evidence of a lack of quality in the estimates and their presence may detract from the credibility of the estimates.

27. Designing a third stage of the weighting methodology which solves these problems is a challenging undertaking. A necessary, although by no means sufficient, condition is that the two sets of control total must themselves be consistent. For example, the number of families times average family size (plus the number of unattached individuals) must be equal to the total number of persons. That condition having been satisfied, it would be necessary to develop an algorithm which further adjusted all of the family and individual weights until equality was restored.

---

[4] Undercoverage is measured as the difference between the sum of the second stage weights and the corresponding control total. In the case of families as the unit of measurement, this measure of undercoverage is a biased measure of underenumeration due to the way the non-response (stage two) adjustment is performed. Where the sample is based on the selection of dwellings with no prior knowledge of the characteristics of the occupants, then the adjustment for non-response results in the weights of all responding household being adjusted by the same amount. In other words, the weights of small households are adjusted by the same amount as the weights for large households. However, small households are more prone to non-response than large households and this means that the non-response adjustment for these households is less than it should theoretically be.

VI. CONCLUSION

28. The family as a social and economic institution is said to be undergoing radical changes in the face of increasing divorce rates, decreasing or chronically low fertility rates, and so forth. Nevertheless, the overwhelming majority of persons continue to live in families and family situations continue to have an influence on the decisions of individuals such as their decision to enter or leave the labour force. In addition, income from employment, or its absence, continues to be the primary determinant of the economic welfare of families and the persons who belong to them. For these and other reasons, the creation of an integrated and harmonized system of family and individual labour market statistics is an increasingly desirable objective.

S-157F
c.2.

MESURES DE L'EMPLOI ET DU CHÔMAGE ORIENTÉES VERS LA FAMILLE

Exposé préparé par

Statistique Canada
Septembre 1983

I. INTRODUCTION

1. Traditionnellement, les données sur l'emploi et le chômage étaient des comptes ou des estimations du nombre de personnes occupées et en chômage. En réalité, chaque personne était considérée comme une entité indépendante et aucune information ou presque n'était publiée sur le contexte familial dans lequel les personnes vivaient, à l'exception peut-être de leur état matrimonial. Cela était peut-être dû en partie au fait que dans beaucoup de pays la source des données sur le chômage était (et est encore dans certains cas) les programmes d'assurance-chômage tandis que celle des données sur l'emploi était (ou est encore) les employeurs. Pour plusieurs raisons, ces deux sources de renseignements sont mal équipées pour fournir des données sur les caractéristiques familiales des personnes en question. L'usage sans cesse croissant des enquêtes auprès des ménages (qu'elles soient ou non les sources des statistiques "officielles") et le fait que les recensements de la population renferment de plus en plus d'informations sur la population active offrent la possibilité d'augmenter considérablement le volume des données permettant de rattacher à sa situation familiale une personne faisant partie de la population active.

2. Le présent exposé traite de certains aspects de la mesure de l'emploi, du chômage et de l'inactivité où les caractéristiques de la famille sont une des dimensions du processus de mesure. La plupart des points considérés concernent l'utilisation des enquêtes-échantillon auprès des ménages comme base pour les estimations.

3. L'exposé commence par une description des deux types généraux de données sur la famille qu'il est possible de générer et propose certaines applications auxquelles ils pourraient convenir. Il se termine par une revue de certains problèmes méthodologiques liés aux statistiques de la population active orientées vers la famille.

II. TYPES DE DONNÉES SUR LA FAMILLE

4. Il y a deux types de base de statistiques de la population active orientées vers la famille. La catégorie la plus habituelle est celle des estimations où la famille est l'unité de mesure. Ces données incluent des estimations comme le nombre de familles biparentales où les deux parents travaillent, le nombre de familles dont un ou plusieurs membres est en chômage et le nombre de personnes seules classées selon leur situation vis-à-vis de l'activité.

5. La seconde catégorie de statistiques prend l'individu comme unité de mesure. Ici, le contexte familial dans lequel vit l'individu sert à caractériser cet individu, au même titre que les caractéristiques habituelles comme l'âge et le sexe. Les statistiques de cette catégorie pourraient inclure des mesures telles que le taux d'activité des femmes ayant des enfants d'âge préscolaire, le nombre de parents célibataires (seuls) en chômage, le nombre de jeunes dans la population active et vivant avec leurs parents, etc.

6. Les deux types de statistiques sont évidemment utiles et le choix de l'un ou de l'autre type dépendra de l'application qu'on veut en faire.

III. CHOIX DU TYPE APPROPRIÉ DE DONNÉES SUR LA FAMILLE

7. Sans sous-estimer les autres avantages que procure un emploi convenable, il faut observer que le revenu a une importance capitale. Bien que le traitement ou salaire soit versé à l'individu, les autres membres de la famille du travailleur

interviennent en général dans la façon dont le revenu est employé. Au Canada, par exemple, 87,2 % des personnes faisant partie de la population active vivent avec au moins une autre personne avec laquelle elles ont un lien de parenté (moyenne annuelle en 1982). Comme le revenu gagné est partagé entre les membres de la famille, la famille est l'unité d'analyse qui convient le mieux quand on veut mesurer l'incidence de l'activité, de l'emploi et du chômage sur le bien-être économique.

8. Par exemple, quand on veut étudier la répartition familiale du revenu, le fait de connaître le nombre de membres d'une même famille qui ont un emploi permet de déterminer plus facilement combien de familles jouissent d'un niveau de revenu "acceptable" du fait que plusieurs de leurs membres ont un revenu et combien de familles parviennent à ce même niveau acceptable avec un seul soutien. Cette possibilité permet elle-même d'évaluer quelle incidence ont sur le bien-être économique des familles les programmes de création d'emplois ou, au contraire, les politiques restrictives qui peuvent avoir pour résultat de réduire l'emploi. Pour prendre un exemple simple, les familles qui perdent le revenu d'un membre dont l'apport au bien-être économique de l'unité familiale est minime souffrent moins que celles qui perdent leur seule source de revenu tiré d'un emploi rémunéré.

9. Ce type de statistiques familiales a également plus d'applications analytiques; il permet entre autres de désagréger la croissance du revenu familial en deux composantes : l'augmentation du revenu des membres de la famille et l'augmentation du nombre de membres de la famille qui ont un emploi.

10. Quand l'unité de mesure est l'individu, on peut relier ces données à toutes les mesures classiques, par exemple le niveau d'emploi et de chômage, et aux mesures dérivées, comme les taux de chômage et les taux d'activité. Plus précisément, on peut utiliser les caractéristiques familiales de l'individu pour désagréger les mesures traditionnelles des niveaux et des taux et ainsi rendre ces mesures classiques plus utiles et augmenter leur "pouvoir explicatif".

11. À titre d'exemple d'un accroissement du "pouvoir explicatif" (ou, dans le cas qui nous occupe, du "pouvoir de prévision"), on peut citer les taux d'activité des femmes, désagrégés selon la présence ou l'absence d'enfants d'âge préscolaire, et les taux d'activité chez les hommes âgés vivant en famille par opposition à ceux qui vivent seuls. Les données canadiennes montrent que la hausse du taux d'activité des femmes est attribuable en grande partie à la hausse toujours plus grande du taux d'activité d'un sous-ensemble, soit celui des femmes ayant des enfants d'âge préscolaire, et que la baisse du taux général d'activité des hommes âgés s'explique essentiellement par une baisse du taux d'activité des hommes qui ne vivent pas avec des personnes avec lesquelles ils ont un lien de parenté.

12. En ce qui concerne la mesure du bien-être économique, il peut être préférable d'utiliser la famille comme unité de mesure et d'expliquer les variations des niveaux de bien-être ou les déterminants du bien-être économique de la famille en fonction du nombre de personnes occupées ou en chômage. Toutefois, la famille n'est pas l'unité de mesure habituellement utilisée pour l'établissement de statistiques sur les niveaux d'emploi ou de chômage. Donc, quand il y a des avantages à recueillir des données en utilisant l'individu comme unité de mesure (et ces avantages peuvent n'être rien de plus que la fidélité à une pratique établie), il est utile de désagréger les données selon la situation familiale lorsqu'on tente d'expliquer la signification des statistiques en termes de bien-être économique. Pour prendre un

exemple simple, une hausse du niveau global du chômage correspondant à un accroissement du nombre des personnes vivant dans des familles dont au moins deux membres sont employés à temps plein a, en termes de bien-être, des conséquences bien différentes de celles d'une hausse de même valeur attribuable en grande partie à des personnes vivant dans des grosses familles dont aucun des membres n'est occupé.

13. Les données sur la famille quand l'individu est l'unité de mesure peuvent également être très utiles pour ce qu'on pourrait appeler des applications analytiques. Pour ne citer qu'un exemple, en périodes de chômage croissant, une augmentation "contre-cyclique" du taux d'activité de certains groupes démographiques peut être attribuée de façon plus précise à l'effet dit du "travailleur supplémentaire" si l'on peut montrer que cette augmentation est causée par des personnes membres de familles dont aucun membre n'est occupé. (Il est certain que les estimations familiales longitudinales sont celles qui conviennent le mieux à ce type d'application, mais une analyse de ce genre de données dépasse les limites du présent exposé). D'autres applications du même genre sont également possibles : par exemple, on peut se servir de ce type de données pour expliquer une baisse des inscriptions dans les maisons d'enseignement postsecondaire en termes du fait que les parents des personnes en question sont occupés ou en chômage.

14. Outre qu'elle nous aide à mieux comprendre les raisons qui motivent la décision d'entrer ou non dans la population active, la dimension familiale peut favoriser une meilleure compréhension du type d'activité. Par exemple, on pourra découvrir que des personnes ont choisis de travailler à temps partiel à cause de la présence d'autres membres de la famille, par exemple de très jeunes enfants ou des personnes très âgées, auxquels elles doivent prodiguer des soins pendant le temps libre que leur laisse l'emploi à temps partiel.

IV. TECHNIQUES DE PRODUCTION DES DONNÉES

15. Les statistiques du marché du travail proviennent de deux sources principales: les sources administratives et les enquêtes auprès des ménages (dont les recensements de la population). Des statistiques administratives, on ne peut rien dire qui serait le reflet universel de pratiques nationales. Cependant, ce ne serait pas trop déformer la réalité que de dire que la plupart des programmes dont sont tirées les statistiques administratives orientées vers le marché du travail produisent des dossiers qui représentent les personnes auxquelles ces programmes s'adressent, qu'elles soient occupées, en chômage ou en dehors de la population active.

16. Dans ces circonstances, il est très difficile d'établir des statistiques familiales, que l'unité de mesure choisie soit l'individu ou la famille. La mesure dans laquelle des statistiques familiales peuvent être produites dépend de la quantité d'information sur la famille que contiennent les dossiers. Même quand une telle information constitue une partie des renseignements sur les participants individuels des programmes, il est peu probable qu'elle soit très complète puisque selon toute vraisemblance elle touchera, de la famille d'une personne, seulement les aspects qui intéressent l'administration du programme duquel les dossiers proviennent. Toute diversification des données sur la famille supposera que les administrateurs des programmes recueillent plus de renseignements des participants aux programmes ou à leur sujet. Si l'on considère les priorités des administrateurs des programmes, on comprendra leur manque d'empressement à se charger de tâches supplémentaires uniquement pour répondre à des objectifs statistiques.

17. Dans le cas des enquêtes auprès des ménages et des recensements de population, les méthodes employées conviennent généralement bien à la production des deux types de statistiques familiales même si dans l'élaboration des méthodes la production de l'un ou l'autre ou des deux types de statistiques n'était pas une considération importante. Les conditions nécessaires sont les suivantes :

i) Toutes les personnes occupant un logement doivent être incluses dans l'enquête.
ii) Un dossier distinct et reconnaissable doit être produit pour chaque membre de la famille.
iii) Les dossiers de l'enquête ont un "code de couplage", de telle sorte que tous les membres d'un ménage ou d'une famille donnée puissent être identifiés dans les fichiers de l'enquête.
iv) La série de questions adressées à tous les répondants doit contenir une question sur la relation (conjoint, fils, fille, etc.) entre les membres de la famille ou du ménage.

18. Ces conditions satisfaites, on peut alors produire les deux types d'estimations familiales en suivant ce qu'on pourrait appeler un processus en quatre étapes:

i) La première étape consiste à rassembler les dossiers de l'enquête en groupes, chaque groupe formant une famille. (Dans le cas des personnes seules, un tel "groupe" consiste en un seul dossier.)
ii) Dans la deuxième étape, on établit les caractéristiques de la famille à partir du groupe des dossiers individuels. Un comptage simple du nombre de dossiers établit la taille de la famille; on peut alors facilement procéder à une désagrégation pour déterminer le nombre de membres de la famille appartenant à chaque catégorie d'âge. Pour les membres d'âge actif, un renvoi à leur situation vis-à-vis de l'activité permet de déterminer combien sont occupés, en chômage ou inactifs. À ce stade, certaines combinaisons peuvent aussi être déterminées. Par exemple, dans les familles époux-épouse, on peut déterminer les cas où l'époux et l'épouse ont un emploi, ceux où un seul des deux conjoints a un emploi, ceux où aucun des deux n'a un emploi et ainsi de suite. Certaines compositions familiales peuvent également faire l'objet de dossiers, par exemple la famille dont une personne est le "chef" (et si cette personne est un homme ou une femme), la famille où sont représentées une, deux ou trois générations, etc. En résumé, il est possible de définir un vecteur plutôt long de caractéristiques familiales, la longueur du vecteur n'étant limitée que par le nombre de caractéristiques de chacun des membres de la famille enregistrées par l'enquête.
iii) Ce vecteur de caractéristiques familiales est ensuite traité de deux façons. D'abord, il est utilisé comme tel comme un nouveau dossier. Le rassemblement de tels dossiers constitue alors le fichier d'où proviennent les estimations ayant la famille comme unité de mesure. Ensuite, ce vecteur est ajouté aux dossiers représentant chacun des membres de la famille ou du ménage en question.
iv) L'étape finale consiste à pondérer les éléments des deux fichiers pour pouvoir convertir les résultats de l'échantillon en estimations des caractéristiques de l'ensemble de la population. Il est évident qu'il faut utiliser de tels poids dans le cas des enquêtes-échantillon, mais ils peuvent également être nécessaires dans le cas des rencensements, du moins quand certaines questions des recensements ne s'adressent

qu'à une partie de la population. (Par exemple, dans le recensement de la population de 1981 au Canada, un ménage sur cinq seulement était interrogé sur sa situation vis-à-vis de l'activité.) La section suivante traite des questions relatives à la pondération des éléments des deux types de fichier.

V. PROBLÈMES D'ESTIMATION

19. On ne compte plus les études qui ont été écrites au fil des ans sur "l'harmonisation" des données provenant de sources différentes. Ces documents traitent souvent de cas où les inconsistences résultent de l'utilisation de méthodes différentes (par exemple, enquêtes auprès des établissements par opposition aux enquêtes auprès des ménages) et des différences dans les erreurs de mesure associées à ces méthodes.

20. Dans le cas des statistiques familiales, des problèmes similaires de compatibilité peuvent survenir même si les deux types de statistiques familiales proviennent de la même source, c'est-à-dire d'une même enquête auprès des ménages. Puisque les deux types d'estimations sont faites à partir du même ensemble de base de dossiers, les disparités ne peuvent provenir que des différences dans les méthodes de pondération.

21. En simplifiant un peu, on peut dire que la pondération se fait en trois étapes. Dans un premier temps, on attribue à chaque dossier un poids initial qui est l'inverse du taux de sondage. Dans le cas des dossiers provenant d'enquête basées sur des échantillons de logements ou ménages, l'inverse du taux de sondage sera le même qu'il s'agisse du dossier d'un individu, d'une famille ou d'un ménage. Par exemple, dans une enquête dont le taux de sondage est d'un ménage sur deux cents, le poids initial sera de 200. La deuxième étape consiste en une série d'ajustements, le plus important étant l'ajustement à faire pour tenir compte des non-réponses. Comme on considère que les non-répondants sont des ménages au complet plutôt que des membres des ménages, l'ajustement pour tenir compte des non-réponses est appliqué uniformément aux dossiers de familles ou d'individus. Quand les méthodes de pondération sont du genre qu'on vient de décrire, les poids sont les mêmes pour toutes les unités de mesure (c'est-à-dire les ménages, les familles et les individus) après les deux premières étapes.

22. Dans certaines enquêtes, la méthode d'estimation ne comporte que ces deux étapes. Dans le contexte qui nous intéresse, cela a l'avantage très réel de préserver la compatibilité des estimations des deux types de données sur la famille. En d'autres mots, le nombre estimé de personnes vivant dans des familles de taille "n" sera exactement égal au nombre de familles de taille "n". De même, le nombre de femmes mariées occupées et vivant avec leur conjoint sera précisément égal au nombre d'hommes mariés occupés et vivant avec leur conjointe. Cette propriété découle du fait que le poids affecté à une famille est égal aux poids associés aux dossiers représentant chacun des membres de la famille en question. Un système de pondération de cette nature présente toutefois des inconvénients sérieux. En particulier, pour un échantillon donné, la variance aléatoire sera plus grande que si l'on avait appliqué des algorithmes de pondération comportant plus d'étapes et les estimations seront caractérisées par un biais vers le bas dû au sous-dénombrement. (Il y a sous-dénombrement quand des logements ou des personnes vivant dans des logements ne sont pas en fait couverts, alors qu'ils auraient dû être dénombrés suivant la conception de l'enquête. Il s'ensuit que le taux de sondage de fait est inférieur aux taux théorique et que la pondération ne compense pas pour cet écart.)

Cette dernière faiblesse est particulièrement sérieuse dans le cas des statistiques sur la famille parce qu'il a été constaté que certaines catégories d'âge et de sexe sont plus sujettes que d'autres à être sous-dénombrées et que les petites familles et les personnes seules sont plus susceptibles de ne pas être dénombrées que les grosses familles.

23. Dans le but de minimiser la variance aléatoire et de faire des ajustements pour tenir compte du sous-dénombrement, on introduit souvent dans les enquêtes-échantillon une troisième étape de pondération qui consiste à ajuster les résultats à des "totaux de contrôle" provenant de sources statistiquement indépendantes de l'enquête.[1],[2] En d'autres mots, les poids affectés sont modifiés de telle sorte qu'une fois la modification apportée, la somme des poids égale les totaux de contrôle. Pour le fichier des individus, on utilise un certain nombre de totaux de contrôle spécifiques pour l'âge et le sexe tandis que pour celui qui représente les unités familiales, les totaux de contrôle sont désagrégés selon la taille et la famille.[3] Les avantages qu'offre cette troisième étape de pondération augmentent, jusqu'à une certaine limite, avec l'acroissement du degré de désagrégation des totaux de contrôle. (Nous n'ananlyserons pas ici la nature de cette limite).

24. Malheureusement, alors que cette troisième étape de pondération offre des avantages tant pour les estimations des caractéristiques familiales des individus que pour les estimations liées aux familles, c'est elle qui introduit l'incompatibilité entre des éléments choisis de l'ensemble des estimations liées à la famille et pour lesquelles l'individu est l'unité de mesure.

---

1 Pour une analyse de l'effet de la troisième étape de pondération sur la variance aléatoire, voir M. Lawes et M.P. Singh, Comparison of Some Ratio Estimators For Large Scale Household Surveys, mémoire présenté à la réunion de 1980 de la Société statistique du Canada.

2 Un autre avantage de cet ajustement est qu'il tend à réduire la variation des estimations d'une observation à l'autre. Cet effet découle de l'absence de variation illusoire dans les totaux de contrôle, qui est généralement une propriété voulue de telles estimations. Alors que cette variation réduite entre les estimations de l'enquête ne présente aucun avantage statistique d'un point de vue purement scientifique, les estimations qui produisent des changements "raisonnables" d'une observation à l'autre (tant en termes de grandeur qu'en termes de signe) sont plus susceptibles d'avoir plus de crédibilité aux yeux de la majorité des utilisateurs de données. Plus les observations sont nombreuses, plus cette propriété sera appréciée.

3 Dans certains cas, au lieu d'utiliser des totaux de contrôle du nombre des familles pour faire une troisième estimation, on peut simplement prendre le poids associé à un membre de la famille, une fois le troisième ajustement fait, et l'affecter à la famille. Cela pourrait être une solution partielle au problème de sous-dénombrement à défaut d'un ensemble de totaux de contrôle du nombre des familles qui conviennent. Cette solution partielle ne résout cependant pas le problème d'incompatibilité dont il a été question à partir du paragraphe 24.

25. Cette incompatibilité découle du fait que le sous-dénombrement varie selon l'âge et le sexe et selon la taille de la famille.[4] La troisième étape de pondération introduit donc des différences dans les poids affectés aux divers membres d'une famille donnée et entraîne l'affectation à une famille d'un poids différent des poids données aux divers membres de la famille en question. Il en découlera des contradictions logiques telles le nombre de conjoints différant du nombre de conjointes, le nombre de "chefs" de famille différant du nombre de familles et le revenu total des familles différant du revenu total des individus.

26. En règle générale, ces écarts ne seront pas très grands dans une enquête bien conçue et bien menée. Quand ils sont grands, cela signifie que le sous-dénombrement est marqué. Le manque de compatibilité ne devrait pas inquiéter les utilisateurs de données qui connaissent les techniques les plus récentes et qui comprennent les méthodes d'estimation et les limites des estimations. Cependant, pour beaucoup d'utilisateurs et peut-être pour la plupart, ces écarts peuvent sembler être un manque de qualité des estimations et leur présence pourrait porter atteinte à la crédibilité des estimations.

27. L'élaboration d'une méthode de pondération de troisième niveau qui résolve ces problèmes n'est pas une tâche facile. Une condition nécessaire mais en aucun cas suffisante est que les deux ensembles de totaux de contrôle soient eux-mêmes compatibles. Par exemple, le nombre de familles multiplié par la taille moyenne des familles (plus le nombre de personnes seules) doit être égal au nombre total de personnes. Une fois cette condition satisfaite, il serait nécessaire de développer un algorithme qui corrige encore plus tous les poids affectés à la famille et à l'individu jusqu'à ce que l'égalité soit rétablie.

VI. CONCLUSION

28. On dit que la famille, en tant qu'institution sociale et économique, subit des transformations radicales du fait que le taux de divorce s'accroît, que le taux de fécondité baisse ou qu'il est chroniquement bas, etc. Néanmoins, la très grande majorité des gens continuent de vivre en famille et la situation familiale continue d'influer sur les décisions personnelles comme celle d'entrer dans la population active ou d'en sortir. De plus, le revenu tiré d'un emploi, ou son absence, continue d'être le déterminant principal du bien-être économique des familles et des personnes qui les composent. Pour ces raisons entre autres, la constitution d'un système intégré et harmonisé de statistiques familiales et individuelles du marché du travail représente un objectif de plus en plus souhaitable.

---

[4] Le sous-dénombrement équivaut à la différence entre la somme des poids de la deuxième étape et le total de contrôle correspondant. Quand la famille est l'unité de mesure, cette mesure du sous-dénombrement est une mesure biaisée du sous-dénombrement dû à la façon dont l'ajustement pour tenir compte des non-réponses (étape deux) a été fait. Dans les cas où on a défini l'échantillon suivant le critère des logements sans connaître au préalable les caractéristiques des occupants, l'ajustement pour tenir compte des non-réponses a alors comme résultat que les poids affectés à tous les ménages qui ont répondu subissent tous le même ajustement quantitatif. En d'autres mots, les poids donnés aux ménages peu nombreux subissent le même ajustement quantitatif que les poids affectés aux ménages nombreux. Cependant, les ménages peut nombreux sont plus susceptibles de ne pas répondre que les ménages nombreux, et cela signifie que l'ajustement pour tenir compte des non-réponses dans la catégorie des ménages peu nombreux est inférieur à ce qu'il devrait être théoriquement.

POOR

QUALITY

ORIGINAL